DOCUMENT RESUME

ED 139 396 IR 004 752

AUTHOR Swift, D.F.; And Others
TITLE Information Systems for Subject Specialists: A Multi-Modal Approach to Indexing and Classification.
PUB DATE Apr 77
NOTE 19p.; Paper presented at the annual meeting of the American Educational Research Association (New York City, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Classification; Educational Sociology; *Indexing; Information Systems; *Specialists; *Subject Index Terms; Systems Development
IDENTIFIERS *Multi Modal Indexing

ABSTRACT

A fundamental problem in the two broad approaches to indexing in the social sciences--providing structure using preferred terms, cross references, and groupings of sets of materials, or compiling a concordance of an author's terms based on occurrence, leaving users free to impose their own structure--is that different indexers and users focus on different aspects of documents as significant. The multi-modal system of indexing describes materials systematically from different viewpoints, handling separately the aspects of documents that come into focus when viewed from each standpoint. Documents are organized according to relatively mutually exclusive categories, described and placed in the dimensions of: (1) theoretical orientation; (2) research method; (3) empirical situation--descriptive or prescriptive; (4) data collection, and (5) form--e.g., bibliography, reader. The system enables users to analyse the search problem by successively partitioning the materials according to dimension and subject headings within the dimension. A feasible application would provide subject specialists with a single tool, more appropriate to their needs, for accessing materials covered by small scale area-specific services. The development of the process is outlined, and an example of indexing provided.
(Author/KP)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

16.20

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED139396

IR004752

Information systems for subject specialists: a multi-modal approach to indexing and classification

Annual Conference of The American Education Research Association, New York City, April 4-8th, 1977

The Open University
United Kingdom
February, 1977

D.F.Swift
V. Winn
D. Bramer

A paper presented to the Symposium on Knowledge Utilization in Educational Administration.

A fundamental problem in indexing and classification in the social sciences is that different people have different ways of 'carving up' the literature. Consequently they see different things in documents. If one were to ask a number of social scientists to invent a title for a given document, they would be likely to respond in varying ways. These variations will not be simply grammatical, but due to the fact that they have chosen to focus on different aspects of the document as the significant ones. Their differing views of documents must all be indexed if they are not to miss relevant material. Established methods of indexing are based on the assumption of a single agreed way of 'carving things up', consequently they are inappropriate in the social sciences.

There are two broad approaches to index organisation. One is to provide structure by which users may locate required material amongst the rest. This is accomplished by designating preferred terms as headings. Further structure may take the form either of cross references in an alphabetic arrangement or physically grouping related sets of material together, and may be imposed a priori or ad hoc. We have called these methods objectivist*. We say this because they are based on a view of the subject matter treated in

*See Swift, D.F., Winn, V.A and Bramer, D.A 'Methodological issues and information research', A Case study in indexing and classification in the sociology of education. Milton Keynes, The Open University, Vol.1, 1973. Also, Watson, L.E. et al 'Sociology and information science', Journal of Librarianship, 1973, 5(4), 270-283.

documents as being self evident. Index structure is devised by 'discovering' all elements of this subject matter and 'mapping' their interrelationships. We would argue, in the spirit of current thinking in the sociology of knowledge, that an index compiler must be recognised to have a particular view of the world, and that there are necessarily many others. The weakness of this approach is that it will lead some users to overlook relevant documents. In 'carving things up' in a particular way, indexers impose structure which will by definition conflict with some views of documents, and the documents will be described in ways which for some users misrepresent them.

The alternative approach is to let documents speak for themselves, as in compiling a concordance of an author's terms, thus leaving the user free to impose his (or her) own structure. This approach appears to involve eclecticism. We say this because documents are grouped into sets on the basis of occurrence of terms, without regard to the point that authors like others have differing ways of 'carving things up'. Documents so grouped will lack coherence as sets. A given set will not contain all the material that a user might expect to find there, and given material may be located in a set labelled by a term which for the user is unexpected if not inappropriate. The weakness of this approach is that indexing lacks predictability. A user must scan the entire index to ensure that he traces all terms by which required material may be labelled, and turn up all documents in sets which he cannot positively eliminate from consideration on the term alone. This search strategy

is practical only in indexes relating to bodies of material that are very narrow in scope or small in quantity.

It is no solution to try to achieve the advantages of both approaches, without the disadvantages of either, by using cross references to help the user to find his way around an unstructured list of terms. If terms are seen as labels for sets of documents, the potential relations between them are infinite. Consequently the cross reference structure can only be arbitrary, and hence there is still a need for end to end scanning. For example, the user who accesses an index via 'social class' may wish to be directed to material on the effects of values, income, occupation, which for him are synonymous with social class. The index cannot possibly anticipate all 'synonyms' for each entry term. The cross referencing will necessarily be selective. The indexer therefore again finds himself imposing a preferred structure on documents.

We set ourselves the task of devising structure which would organise data about documents in a predictable way, yet without imposing serious limits on the scope of material that can be handled effectively, and without interfering seriously with the ways in which users may wish to approach the literature. We believe that we have produced a workable solution to this problem. Crudely, the principle is that we describe material systematically from a range of different standpoints, which are then made explicit in the organisation of the index. The point of organising the index according to standpoint is that documents are interrelated

in complex ways. By handling separately the aspects of documents that come into focus when they are viewed from a given standpoint, we can devise a coherent structure for that way of viewing documents. This means that the user, having chosen a standpoint, can make predictions from a given set of material as to the other sets in which required material will be located if it is in the system.

In theoretical terms, we see ourselves as having rejected an objectivist position. At the same time, we avoid a slide into relativism by adopting what may be called a pluralist position. Those holding an objectivist view of knowledge will tend to assume that denial of the self evident nature of document subject matter entails operating in terms of users' idiosyncratic views of the world. The reductio ad absurdum criticism of our position would be to expect that our solution must be one user/one index. This does not follow. We make the assumption that views of the world are analysable on common dimensions. Analogously with Kuhn's notion of paradigm, we think in terms of dimensions on which users commonly differentiate amongst their material (eg problem focus, theoretical stance). These dimensions represent a framework by which an indexing system can generate and organize data about documents in ways which have relevance for users. Our solution is best described as a 'multi-modal' approach.

Information scientists may feel that the multi-modal approach throws the baby out with the bathwater. Our dimensions are clearly not mutually exclusive - for example, the theoretical

stance a person chooses to adopt defines what for him counts as a problem. In these circumstances, indexing would seem likely to lack reliability, with consequent difficulty, for a user trying to locate required material, as great as any of the approaches we have just discussed. However, by keeping the descriptive data associated with each dimension separate, we believe that documents can be organised according to relatively mutually exclusive categories. This enables the indexer to make reliable decisions concerning the placement of documents - a necessary condition for predictibility in use.

At the same time, however, the structure of a multi modal index is accordingly more complex than that of indexes compiled by established methods. Users may feel that its complexity involves greater effort in use than other types of index. Most users would ideally like an index which can be used in a mechanical way - think of a word, look it up, the material you find is the material you want. This is the mode of use assumed by conventional indexes, but they are rarely as simple to use as that, at least if the user takes the necessary steps to ensure that he does not overlook relevant material. The effectiveness of such indexes can be enhanced by restricting the scope of the documents covered to those of interest to people who share a common view of the world. Many users would find an index of this kind adequate for their practical purposes, and would prefer its simplicity in use. Complexity of structure only becomes necessary because we are bringing diverse material together - a practice we adopt because we are seeking to cater for

users' needs, even though this might conflict with their immediate preferences. We all recognise, although we may not live by it, that different views of the world have much to say to one another, and that this is a means by which ideas develop and change. Moreover, provided that structure is meaningful, greater complexity need not seriously inconvenience the user and, in fact, probably involves less effort in use for a given quality of outcome.

Operationalising the multi-modal approach

The dimensions we have selected allow us to describe and group documents in terms of theoretical orientation, method of research, empirical situation under study (concern may be either description, prescription), data collected and form (eg bibliography, reader). The indexer systematically takes these as viewpoints and generates a set of title-like phrases for each document, each phrase taking a different aspect of document content as its focus but referring to the whole. The phrases associated with each dimension are presented as separate series of data (in effect as separate indexes), each organised according to its own set of headings. Thus pieces of research using a common theoretical orientation, for example, are grouped together, and the further information about the individual items enables the user to select from them according to his particular purpose.

An example will best indicate the nature of the approach. The following abstract outlines a document recently indexed:

261 Rothstein, Stanley William (California State U of Fullerton) Researching the power structure: personalized power and institutionalized charisma in the principalship. INTERCHANGE, 1975,6(2), 41-48.

This paper concerns itself primarily with an exploration of aspects of the structure and interaction in an urban junior high school as they affected the decision making processes at the school level. The concern is with the divergence that exists between the apparent bureaucratic rationale of urban education and the relational elements that actually govern such institutions. The goal is to demystify the current tendency in educational literature to view schools as bureaucratic entities. Instead, evidence points to the office of the principalship as a focus of institutionalised charisma and 'bossism' in our society. The specific data of the study describes the author's experience in one ghetto junior high school and concern the events that students and staff participated in during 1966-1970. A number of 'decisions' made in the school are analysed, decisions which showed the school to be functioning as a custodial institution with educational pretensions and the administrators and staff to be unable to construct situational definitions for learning. The decision making processes in this ghetto school were arbitrary. The principalship was the center of a highly personalised power system and an office best understood as a focus of institutionalised charisma in our society. Planned educational change will have to reckon with the reality of this highly personalised power structure that has developed around the office of the principal. Only in-depth studies of actual school situations will be able to provide educational planners with the essence of arbitrary rule that underlines the appearance of bureaucratic procedure in public education.

In one series of data, focussing on the situation under study, the document may be described:

> Power structure and the office of principal in schools: a study of decision making in a ghetto junior high school, with implications for educational planners

In this context, the document forms a member of sets such as Administration and decision making, Political structures and processes, Teachers and teaching profession, Secondary education.

In another series of data, focussing on the way in which the author conceptualises his subject matter, the document may alternatively be indexed by the major concepts employed - Power and Charisma, then being described:

> Principalship as focus of personalised power and institutionalised charisma: interpretation of study of power structure in one ghetto junior high school

The entry might also appear amongst those for studies employing both structural and interactional approaches.

Yet again, in terms of the data collected for this study, the document would be brought together with others relating to ghetto schools.

The value of the approach is that it enables us for example to include in our index both a set of documents which conceptualise situations in terms of power and a set of documents which analyse actual power structures. There is some overlap between these sets but membership of one does not necessarily imply membership of the other. In conventional indexing, these sets would either be merged, with the problem for cross referencing that we have described, which arises because different organising principle are being used simultaneously. Alternatively, the indexer would select one of these organising principles and ignore all others, with consequent risk of information loss, because the user who wishes to search in terms of a non-preferred organising principle will not be able to do so. Similarly, some but not all studies of secondary schools are concerned with ghetto schools, and some but not all studies of ghetto

schools are generalisable to the whole population of secondary schools.

Research leading to formulation of multi-modal approach

The multi-modal approach was first conceived of as a means of coping with the problems of indexing the literature of the sociology of education, which is an area of study marked by conflicts of approach so sharp that its members tend to 'talk through' each other. One of the activities of the group is to publish Sociology of Education Abstracts and, as the material in the system began to accumulate, we were anxious to make it accessible to retrospective searching. We therefore experimented with a range of the available indexing systems, alphabetical and classified, pre-and postcoordinate (including the new British PRECIS system and an Anglicised version of ERIC indexing). The idea was to select the one which worked best for sociologists of education.

Users evaluated our sample indexes but, although they were able to give us a rank order of effectiveness, the more important point to emerge was that they felt happy with none.

They accepted that some sort of structure must be imposed if they were to be able to locate their material in amongst the rest but the kind of structure imposed by the existing schemes hindered as much as helped in this respect*. The

*Services for social scientists in general are greatly underused. (Line, M.B. et al Information requirements of researchers in the social sciences. Bath, Bath University Library, 2 vols., 1971), and this is arguably one of the reasons.

main difficulty was that relevant documents were inadequately described. It was not clear quite what was lacking, but the result was that the description gave a misleading impression of documents - a situation which did not inspire confidence in the indexing.

Seeing no alternative at the time, we selected PRECIS* as one of two alphabetical pre-coordinate indexes which came almost equal first, and experimented further to find out what was going wrong and how to adapt the system to our needs. The key to the PRECIS system lies in a set of special procedures by which statements of document subjects are formulated. The indexer prepares a title-like phrase (eg Role of the principal in decision making in a ghetto junior high school - to continue with the example used earlier). He (or she) analyses the phrase into different types of elements, which are then set down as a string of terms in a prescribed order. The strings are then rotated to produce a set of entries for the document in question, with all significant terms in a string taken in turn as entry point.

Our experience in using the system was that the procedures could handle a phrase like the one in the preceding paragraph, but were unable to accommodate further information such as that principalship was viewed in terms of the institutionalisation of charisma. As framed, the procedures

---

See eg Austin,D. 'The Development of PRECIS: a theoretical and technical history', Journal of Documentation, 1974, 30(1), 47-102.

thus force the indexer to focus on certain aspects of document content to the exclusion of others, which in some searches would be more crucial in predicting the nature of the material described by given indexing data, and deciding whether to turn it up. We were able to reformulate the procedures in a less restrictive way, but this only led us into a further problem. The overall structure to which indexing is geared ignores prior distinctions amongst different aspects of document content, and groups entries simply on the basis of terms in strings. This practice thus makes it impossible to bring together many documents which we think users would wish to find brought together as sets. This was quite unacceptable.

In this respect, the PRECIS system follows standard indexing practice, and is typical at least of those used in large services. The basic problem is that documents are grouped simply in terms of what they are said to be 'about', whereas other criteria are of equal and sometimes prior importance in users' thinking. For example, in the Current Index to Journals in Education, an article entitled 'The Social organization of learning: initiation rituals and public schools' was indexed under Learning processes, Public schools, Rural schools, Social history, Liberia. The index failed to bring out the reasons why all the schools were failing to educate pupils, namely that schools are organised for indoctrination and for ritual initiation of pupils into a new society rather than for education. Similarly, in the case of the principalship example, the document could not be accessed via the notion of charisma.

We came to the conclusion that a radically different approach was needed. The approach outlined in the previous section was the outcome. We formulated our approach by working in cooperation with subject specialists to produce an experimental index to a sample of the journal literature selected for relevance to the sociology of education.

Implementing the multi-modal approach in practical contexts

Further developmental work was required before we had a scheme which we could use in Sociology of Education Abstracts . In addition, we saw the possibility of developing a system which would handle the whole field of educational studies, rather than taking the sociology of education in isolation from other subdisciplines and specialisms comprising the education field. In this wider context, we had in mind indexing the teaching materials put together by the Open University Faculty of Educational Studies. The Open University is a nationwide institution concerned with distance education of adults by means of multi-media materials. Teaching materials comprise correspondence texts, readings in books and journals, along with associated radio and television programmes. The education faculty output consists of a range of courses produced by different groups of academics, some being discipline based (including the sociology of education), others problem oriented (eg urban education). Our more recent work thus represents both a development from and an extension of our original intentions, in that we are now thinking in transdisciplinary as well as in single discipline terms.

The Faculty materials provided a practical context for selecting dimensions which are discipline relevant yet non discipline specific, for use in generating and organising data. It also enabled experimentation with ways of describing differently structured material, different levels of treatment, and different ways of formatting data. The various options we identified were then tested out on Sociology of Education Abstracts material. As a result, the group now has a working system (though one still under development), comprising procedures for describing individual items, and structure for relating them together. We are currently mapping different bodies of material on to this structure, which involves constructing sets of headings by which to group documents on the different dimensions.

The system now caters relatively adequately for the sociology of education. Other fields of educational studies are as yet treated at a more generalised level. The work so far however appears to have demonstrated the practicability of developing a scheme applicable to the wider field without losing the capability of describing particular material broadly in its own terms and grouping it accordingly. A drawback of the system is that to avoid undue overlap the structure associated with some of the dimensions is necessarily coarse rather than fine. One means of overcoming this is to incorporate other headings, drawing on specialist schemes such as that used in Educational Administration Abstracts. The system would also be complemented by an alphabetic list of the specific terms used in descriptive phrases, with references to the location of the phrases. The current stage of development of the system is illustrated by the sample index in the appendix.

The system in use

Typically, a user comes to an index with the idea that he wants 'everything on X' (eg community control, secondary schools, ethnomethodology). In practice he is implicitly making at least two kinds of qualification. The first is that he means material whose authors treat what he understands by the notion X. The second is that he means material whose authors treat X in ways appropriate to his task. In other words, the typical user does not want everything that someone somewhere might characterise as being about X, and he may well also want some material which others would not characterise as such. He wants to select material according to his particular relevances and priorities.

It is not the job of an index to enter into the nuances of different definitions of terms. Its task is rather to bring together documents similar in content. The user must decide for himself whether one author's notion of X corresponds sufficiently closely to his own to be said to be 'about X', or if what he means by X is what another author calls Y. However, indexing can help on the problem of appropriate treatment. The dimensions first enable him to eliminate 'non-appropriate treatment' material, thus giving him a much smaller body of material on which to focus. The material which remains is that which meets the user's requirements as to treatment. Broad grouping of this material can then enable him to eliminate further material which he would not consider to be about X. This will leave him with a relatively small number of sets of material to be studied in detail for correspondence with what he chooses to mean by X.

The multi-modal type of system helps the user to analyse an initial very generalised statement of his search problem, and by this means to clarify what he 'really' wants. The dimensions may be taken as a set of prior questions to be answered (do I want this or this 'kind of thing'). The headings on a chosen dimension then represent further options from which to choose, whilst the information given about individual documents in any chosen set provides a basis for more detailed decisions as to relevance, as in any index. In short, in a multi-modal system, the search process is one of successive partitioning of the material in the system, 'funnelling down' to a required set.

Looking now at the multi-modal system as it concerns the producer rather than the consumer, the indexer goes through a mirror image of this process. He must provide data about documents relevant to the kinds of decisions users may wish to make in searching the literature and the kinds of criteria they may wish to employ. Thus, he asks prior questions of the following kind: what is the theoretical content of this document if any (eg selection as social control)? what empirical situation is being analysed if any (eg college selectivity effects)? does the document offer prescriptions of any kind, and if so what (eg pros and cons of open admissions)? He then goes on to locate documents under the headings appropriate to them when viewed from each of these standpoints, and the phrase used in each context is constructed so as to include the kind of further information which most effectively distinguishes them from similar material.

The place of the multi-modal type of system in the existing provision of information for education

Having described the multi-modal type of system and some of the research leading to its development, it remains to be shown that there is a need for our system. It was a natural development in our work, in attempting to devise a means of serving the needs of specialists, to widen its scope to the broader field of education. The nature of the multi-modal approach makes it as appropriate to a multi-disciplinary field as to use within a discipline, and the nature of the field is such as to make our approach as necessary, if not more so. Provided that the widening of scope did not preclude treating any given area in its own terms, it might be seen as an obviously desirable next step.

Unfortunately, the inertia of large capital investment being what it is, we cannot envisage a situation in which a multi-modal service would be set up in competition with giants like ERIC. Thus, the main application of the multi-modal approach would seem to be in small scale area-specific services. In this case small scale area-specific systems would seem more appropriate than an education-wide one.

The justification for developing an education-wide system lies for us in a feeling of dissatisfaction with the present pattern of provision for specialists, which involves a variety of separate and differently organised services. Boundaries amongst established schools and specialisms are far from agreed, new and overlapping ones are frequently emerging. At a practical level, gaps remain unfilled, and

at the same time there is much duplication of work. More seriously, there are many users whose interests are not identifiable in terms of the available services, and the situation precludes searching across different bodies of material in any systematic fashion. The larger scale services provide an alternative means of access to the literature but, as we have argued, they are not effective for specialists' purposes.* Some of the smaller scale services were initiated by subject experts for this very reason.

We have come to feel that subject experts and information experts have, in differing ways, taken too simplistic a view of the problems of ensuring effective access to social science material in general and education material in particular. Clearly, the answer is not to be found in any kind of mindless rationalisation or externally imposed order. The burden of this paper is that it is possible to handle a broad range of documents in a which which is integrated yet sensitive to the differing approaches users may wish to make to them. The future we would like to see is one of common indexes to groups of small scale services, for instance all those serving educational studies. The idea would be to provide users with a single tool for accessing the material covered by the smaller services, yet one more appropriate to their needs as subject specialists than the larger ones. The multi-modal system we are developing would be a possible means, though not the only one, to this end.

---

*This paper has confined itself to the problems of indexing and classification. In a wider context there would be much to say about the problems created by lack of quality control at the selection and abstracting stages.